

City University of New York (CUNY)

CUNY Academic Works

Publications and Research

New York City College of Technology

2023

Understanding Data Mining and Its Relation to Information Systems

Malak Alammari

CUNY New York City College of Technology

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/ny_pubs/1044

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).

Contact: AcademicWorks@cuny.edu



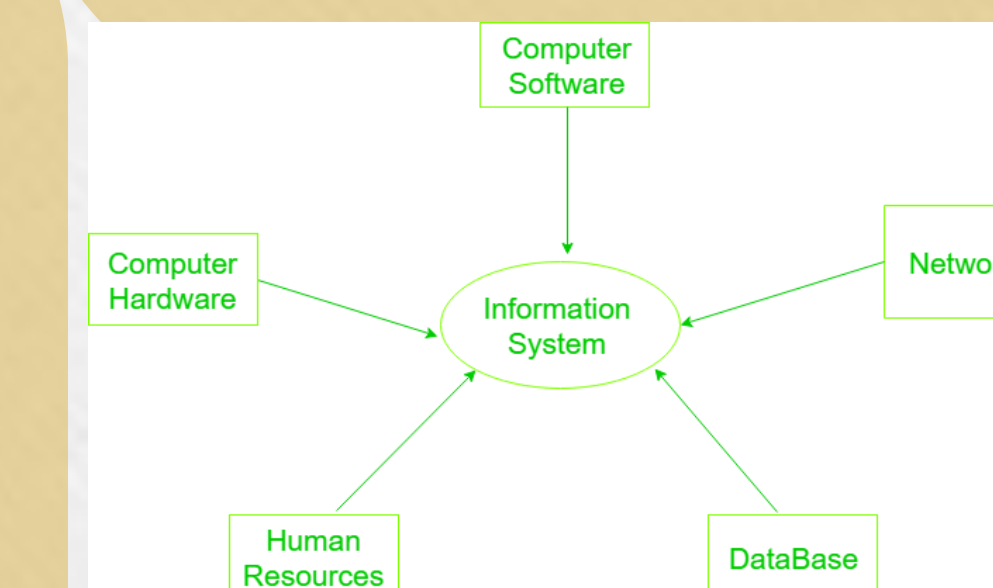
Understanding Data Mining and Its Relation to Information Systems

Author Malak Alammari

Mentor – Professor Patrick Slattery

Components Of Information System

It is the combination of hardware and software that work together to collect and distribute useful data. The database is software that runs on hardware. Hardware job is to store data and software processing data. Information systems were in the middle and everything else is connected to it like Computer Hardware, Computer Software, Databases, Network and Human Resources. Its objective is to provide information, gather the data, process the data and communicate information to the user. Computer Hardware used for input, output and processing. Computer Software is used to control and coordinate the hardware components and it has three types of classified System Software, Application Software and Procedure. Databases are data that is raw facts. They are unorganized and later processed to generate information. Networks include communication media, and Network Support. Human Resources It requires the user to manage the system.



Abstract

This research project aims to enrich an Open Educational Resource (OER) textbook on Introduction to Information Systems/Technology with a focus on data mining and its relation to hardware and software components of information systems. The study will address the following research questions: (1) What is data mining? and (2) How does data relate to the hardware and software components of information systems?

To answer these questions, the researcher will conduct research to ascertain the current state of data mining and its relevance in the field of information systems/technology. The results of the research will be incorporated into an existing OER textbook, modernizing or updating its content. The researcher will collaborate with the project team to develop a plan for incorporating new content into the OER textbook and communicate findings and progress to ensure the project's overall success.

The different type of data mining algorithm

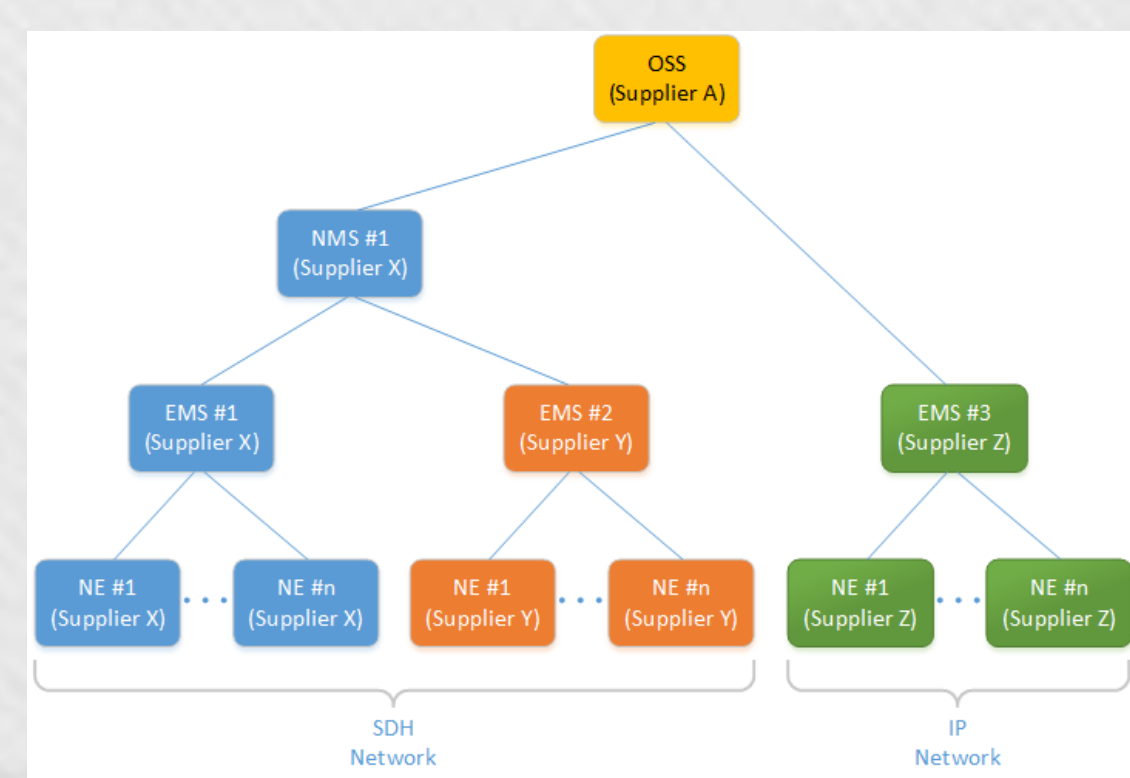
K-Nearest Neighbor -It is one of the most basic yet essential classification algorithms in Machine Learning. The way it works is by storing all the available cases and classifies the new data or case based on a similarity measure.

For example, if you see the graph it has red Diamond shapes and green square. Both are classified but the white triangle is not. We can assign it to a group by observing what group its nearest Neighbors belong to. This means a point close to a cluster of points classified as 'Red' has a higher probability of getting classified as 'Red'.

Neural Network. It performs tasks by being exposed to various datasets and examples without any task-specific rules. It is based on computational models for threshold logic. Neural has three basic sequences of learning. It first is simulated to new environments then the free parameter is going to change. The neural responds in a new way because of the new environment. There are seven types of Neural Networks, an example is Multilayer Perceptron.

Deciding Tree Algorithms- Deciding Tree is the best now for it accurately predicting the outcome and the most popular tool classification. A Decision tree is designed like the roots of tree. It works by node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label.

They are many more different Algorithm



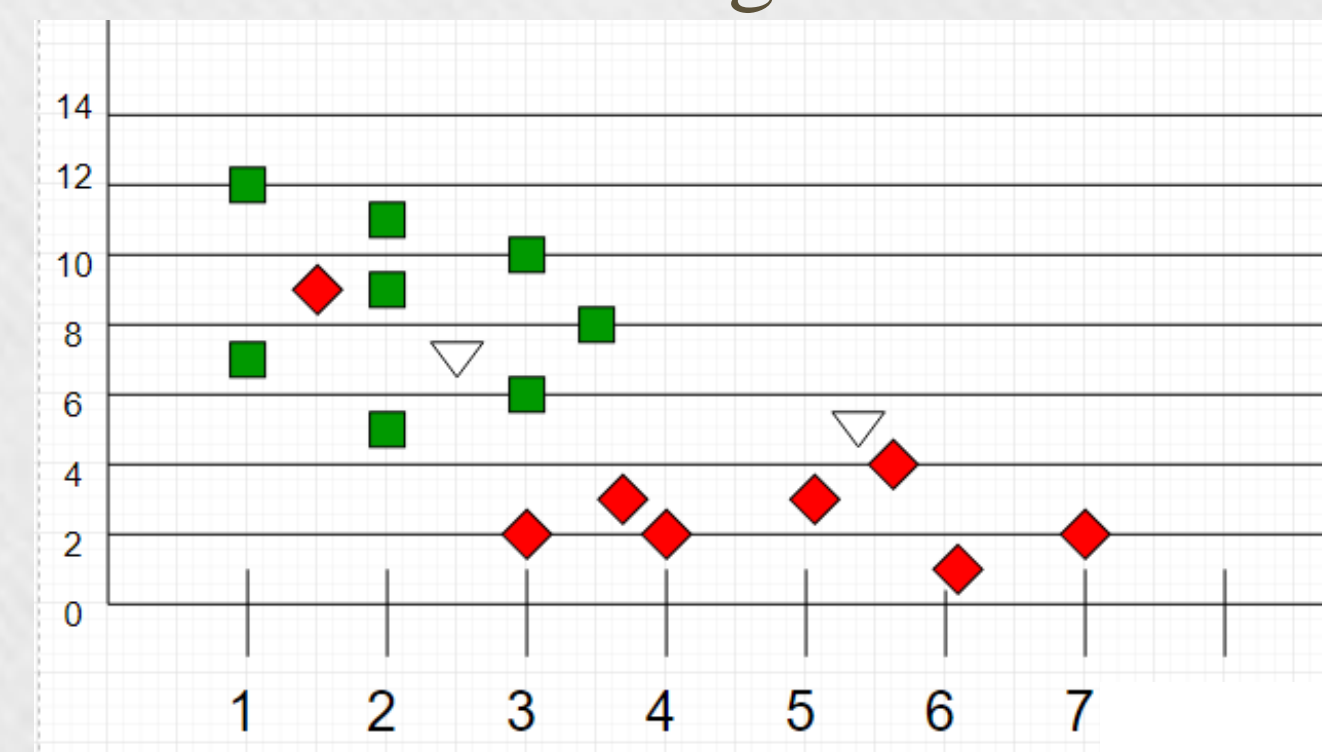
Introduction

Data mining is a process of finding patterns and trends by extracting from large data sets using machine learning and statistical analysis techniques. The idea of data mining has been with us since 1763 when Thomas Bayes published a paper about the Bayes' theorem. Which theorem for relating the current probability to prior probabilities. It isn't the 1990s when the term "data mining" started to appear and become as we know it now.

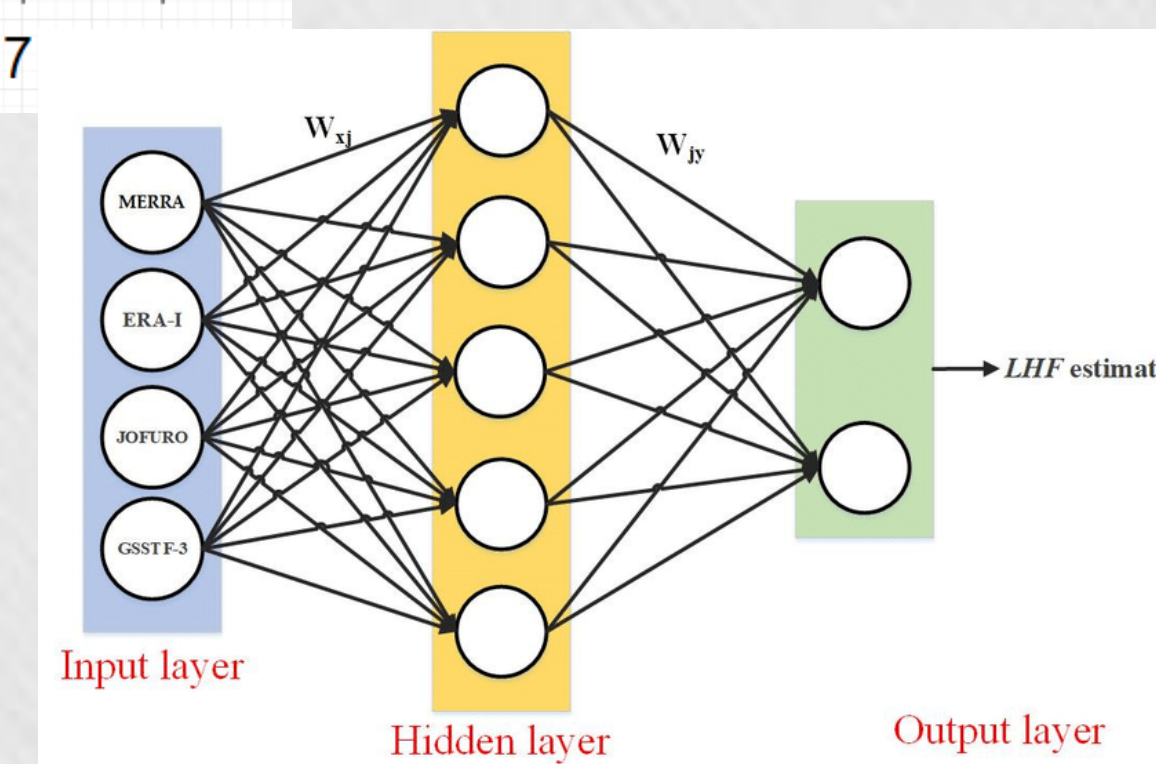
Algorithm

The algorithm first analyzes the data you provide, and it looks for specific types of patterns. It uses the results of analysis over many iterations to find the optimal parameters for creating the mining model. The model can have many different outcomes based on the data you provide. The model can be predictor of the outcome, mathematical model and it can be set of clusters that describes the case in the dataset. When choosing an algorithm is important to choose the right one for the task because each algorithm produces a different result. There are five types of Classification algorithms, Regression algorithms, Segmentation algorithms, Association algorithms and Sequence analysis algorithm.

K-Nearest Neighbor



Neural network



Deciding Tree

Conclusion / Discussion

Data mining is very significant because it can help with product Development because companies can use data to analyze which audience would be the best for the product. It can also help with manufacturing data mining. It can help companies to calculate the raw materials that the product needs, its cost and what would be the best option, that cost the least.

References

-GeeksforGeeks. (2023, April 18). K-nearest neighbours. GeeksforGeeks. Retrieved April 27, 2023, from <https://www.geeksforgeeks.org/k-nearest-neighbours/>
-GeeksforGeeks. (2023, April 18). Neural networks: A beginners guide. GeeksforGeeks. Retrieved April 27, 2023, from <https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/>
-GeeksforGeeks. (2023, April 18). Decision tree. GeeksforGeeks. Retrieved April 27, 2023, from <https://www.geeksforgeeks.org/decision-tree/>
- Minewiskan. (n.d.). Data Mining Algorithms (Analysis Services - data mining). Microsoft Learn. Retrieved April 27, 2023, from <https://learn.microsoft.com/en-us/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=asallproducts-allversions>